

阿里云 E-MapReduce

常见错误排除

文档版本：20190905

法律声明

阿里云提醒您在使用或阅读本文档之前仔细阅读、充分理解本法律声明各条款的内容。如果您阅读或使用本文档，您的阅读或使用行为将被视为对本声明全部内容的认可。

1. 您应当通过阿里云网站或阿里云提供的其他授权通道下载、获取本文档，且仅能用于自身的合法合规的业务活动。本文档的内容视为阿里云的保密信息，您应当严格遵守保密义务；未经阿里云事先书面同意，您不得向任何第三方披露本手册内容或提供给任何第三方使用。
2. 未经阿里云事先书面许可，任何单位、公司或个人不得擅自摘抄、翻译、复制本文档内容的部分或全部，不得以任何方式或途径进行传播和宣传。
3. 由于产品版本升级、调整或其他原因，本文档内容有可能变更。阿里云保留在没有任何通知或者提示下对本文档的内容进行修改的权利，并在阿里云授权通道中不时发布更新后的用户文档。您应当实时关注用户文档的版本变更并通过阿里云授权渠道下载、获取最新版的用户文档。
4. 本文档仅作为用户使用阿里云产品及服务的参考性指引，阿里云以产品及服务的“现状”、“有缺陷”和“当前功能”的状态提供本文档。阿里云在现有技术的基础上尽最大努力提供相应的介绍及操作指引，但阿里云在此明确声明对本文档内容的准确性、完整性、适用性、可靠性等不作任何明示或暗示的保证。任何单位、公司或个人因为下载、使用或信赖本文档而发生任何差错或经济损失的，阿里云不承担任何法律责任。在任何情况下，阿里云均不对任何间接性、后果性、惩戒性、偶然性、特殊性或刑罚性的损害，包括用户使用或信赖本文档而遭受的利润损失，承担责任（即使阿里云已被告知该等损失的可能性）。
5. 阿里云网站上所有内容，包括但不限于著作、产品、图片、档案、资讯、资料、网站架构、网站画面的安排、网页设计，均由阿里云和/或其关联公司依法拥有其知识产权，包括但不限于商标权、专利权、著作权、商业秘密等。非经阿里云和/或其关联公司书面同意，任何人不得擅自使用、修改、复制、公开传播、改变、散布、发行或公开发表阿里云网站、产品程序或内容。此外，未经阿里云事先书面同意，任何人不得为了任何营销、广告、促销或其他目的使用、公布或复制阿里云的名称（包括但不限于单独为或以组合形式包含“阿里云”、“Aliyun”、“万网”等阿里云和/或其关联公司品牌，上述品牌的附属标志及图案或任何类似公司名称、商号、商标、产品或服务名称、域名、图案标示、标志、标识或通过特定描述使第三方能够识别阿里云和/或其关联公司）。
6. 如若发现本文档存在任何错误，请与阿里云取得直接联系。

通用约定

格式	说明	样例
	该类警示信息将导致系统重大变更甚至故障，或者导致人身伤害等结果。	 禁止： 重置操作将丢失用户配置数据。
	该类警示信息可能导致系统重大变更甚至故障，或者导致人身伤害等结果。	 警告： 重启操作将导致业务中断，恢复业务所需时间约10分钟。
	用于补充说明、最佳实践、窍门等，不是用户必须了解的内容。	 说明： 您也可以通过按Ctrl + A选中全部文件。
>	多级菜单递进。	设置 > 网络 > 设置网络类型
粗体	表示按键、菜单、页面名称等UI元素。	单击 确定 。
<code>courier</code> 字体	命令。	执行 <code>cd /d C:/windows</code> 命令，进入Windows系统文件夹。
<code>##</code>	表示参数、变量。	<code>bae log list --instanceid</code> <code>Instance_ID</code>
<code>[]</code> 或者 <code>[a b]</code>	表示可选项，至多选择一个。	<code>ipconfig [-all -t]</code>
<code>{ }</code> 或者 <code>{a b}</code>	表示必选项，至多选择一个。	<code>swich {stand slave}</code>

目录

法律声明.....	I
通用约定.....	I
1 附录.....	1
1.1 集群服务管理页面.....	1
1.2 集群端口配置.....	2
1.3 E-MapReduce 版本相关.....	4
1.4 错误代码表.....	4
1.5 状态表.....	5
1.6 E-MapReduce常用文件路径.....	7

1 附录

1.1 集群服务管理页面

您可以购买一台 1C2GB 配置的 Ubuntu 系统的 ECS 机器放在内网环境中，作为集群的管理页面访问使用。

集群中的所有服务访问地址一览

软件	服务	访问地址
Hadoop		
	yarn resourcemanager	k masternode1_private_ip:8088,masternode2_private_ip:8088
	jobhistory	masternode1_private_ip:19888
	timeline server	masternode1_private_ip:8188
	hdfs	masternode1_private_ip:50070,masternode2_private_ip:50070
Spark		
	spark ui	masternode1_private_ip:4040
	history	masternode1_private_ip:18080
Tez		
	tez-ui	masternode1_private_ip:8090/tez-ui2
Hue		
	hue	masternode1_private_ip:8888
Zeppelin		
	zeppelin	masternode1_private_ip:8080

软件	服务	访问地址
Hbase		
	hbase	masternode1_private_ip: 16010
Presto		
	presto	masternode1_private_ip: 9090
Oozie		
	oozie	masternode1_private_ip: 11000
Ganglia		
	ganglia	masternode1_private_ip: 8085/ganglia

1.2 集群端口配置

Hadoop HDFS

服务	端口	配置	描述
NameNode	9000	fs.default.name 或 者 fs.defaultFS	fs.default.name 已 经过期但仍能使用
NameNode	50070	dfs.http.address 或 者 dfs.namenode. http-address	dfs.http.address 已 过期但仍能使用

Hadoop YARN (MRv2)

服务	端口	配置
JobHistory Server	10020	mapreduce.jobhistory. address
JobHistory Server	19888	mapreduce.jobhistory. webapp.address
ResourceManager	8025	yarn.resourcemanager. resource-tracker.address
ResourceManager	8032	yarn.resourcemanager. address

服务	端口	配置
ResourceManager	8030	yarn.resourcemanager.scheduler.address
ResourceManager	8088	yarn.resourcemanager.webapp.address

Hadoop MapReduce (MRv1)

服务	端口	配置
JobTracker	8021	mapreduce.jobtracker.address

Hadoop HBase

服务	端口	配置
HMaster	16000	hbase.master.port
HMaster	16010	hbase.master.info.port
HRegionServer	16020	hbase.regionserver.port
HRegionServer	16030	hbase.regionserver.info.port
ThriftServer	9099	-

Hadoop Spark

服务	端口	配置
SparkHistory	18080	-

Storm

服务	端口	配置
Storm UI	9999	ui_port

Druid

服务	端口	配置
overlord	18090	overlord.runtime -> druid.port
coordinator	18081	coordinator.runtime -> druid.port

服务	端口	配置
middleManager	18091	middleManager.runtime -> druid.port
historical	18083	historical.runtime -> druid.port

1.3 E-MapReduce 版本相关

- E-MapReduce 的版本会定期升级，类似 1.0、1.1、1.2。
- 每一个版本上捆绑的软件和软件的版本都是固定的。目前还不支持软件的多个不同版本的选择，也不推荐用户自行更改软件的版本。例如我们在 1.0 中内置的是 Hadoop 2.6.0，Spark 1.4.1 等。
- 一旦选择了某个版本，并创建了集群，该集群使用的版本就不会自动升级了，例如 1.0 版本，那么 Hadoop 就一直保持在 2.6.0，Spark 就一直保持在 1.4.1。后续如果版本升级到了 1.1，Hadoop 到了 2.7.0，Spark 到了 1.5.0 也不会影响到已经创建出来的集群。只有新的集群才会使用新的镜像。
- 当您升级集群版本的时候，比如从 1.0 升级到 1.1，防止一些升级的不兼容，环境的变化等，请务必测试您的作业，保证他们在新的软件环境中也能正常的运行。

1.4 错误代码表

常见错误码

错误码	说明
4001	请求参数不符合要求，如参数缺少，参数格式不满足要求
4005	没有权限访问其他用户的资源
4006	该集群处于异常状态，不能提交作业，请您检查执行计划关联的集群是否已经释放
4007	安全组名称不能为空
4009	请检查您的账户是否欠费或者被冻结
4011	该集群处于异常状态，不能恢复调度，请您检查执行计划关联的集群是否已经释放
5012	用户创建的安全组数量超过上限，您可以前往 安全组页面 删除一些无用的安全组

错误码	说明
5038	作业在运行或待运行的执行计划中，暂不能修改，需等待相关联的执行计划运行完成才能修改，您可以克隆一个新的作业进行修改使用
5039	集群角色锁定失败，E-MapReduce 的运行需要一些权限，关于 E-MapReduce 的 角色授权 ，请前往进行 创建
5050	访问数据库错误,请稍后重试
6002	更新集群状态失败
8002	没有操作的权限，请使用 RAM 授权
8003	没有 PassRole 操作的权限，请使用 RAM 授权
9006	集群 ID 不存在，请确认
9007	创建集群时设置 Master 节点的登陆密码，必须包含大小写字母以及数字，长度为 8~30

ECS 相关错误

错误码	说明
指定的 InstanceType 未授权使用	创建集群使用的机型尚未申请。请在 ECS 的 购买页面提交申请开通高配 。
Zone 或者 Cluster 的库存不够了	ECS 在该区域的库存不足

1.5 状态表

集群状态表



说明:

在集群列表和集群详情页面会查看到集群状态。

状态名	状态代码	状态说明
创建中	CREATING	集群正在构建，包括两个阶段：一是物理 ECS 机器的创建，二是 Spark 集群的启动，稍等片刻即可达到运行中状态

状态名	状态代码	状态说明
创建失败	CREATE_FAILED	创建过程中遇到异常，已经创建的 ECS 机器会自动回滚，在集群列表页面单击状态右边的问号，可查看异常明细
运行中	RUNNING	计算集群处于正常运行状态
空闲	IDLE	集群目前没有运行执行计划
释放中	RELEASING	单击集群状态列表的释放按钮可达到此状态，此状态表示集群正在努力释放，用户稍等片刻，即可达到已释放状态
释放失败	RELEASE_FAILED	释放集群过程中出现异常，在集群列表页面单击状态右边的问号，可查看异常明细，出现此状态需要用户重新单击释放按钮
已释放	RELEASED	计算集群以及托管计算集群的 ECS 机器均处于释放状态
状态异常	ABNORMAL	当计算集群中的一个或多个计算节点出现不可恢复的错误时，则集群处于此状态，用户可单击释放集群按钮，释放该集群

作业状态列表



说明:

在作业状态列表里会查看到作业状态

状态名	状态说明
任务就绪	作业创建信息完整正确，并被成功的保存，正在准备提交到系统的调度队里中，稍等片刻会处于提交中状态
提交中	作业在计算集群的提交队列中排序等待，还未提交给计算集群进行计算
提交失败	作业在提交到计算集群的过程中，发生异常，如需再次执行该作业，用户需要克隆作业重新提交

状态名	状态说明
运行中	作业正在计算集群中拼命的运算，请稍等片刻，单击作业列表中相应的日志按钮可实时查看输出日志
运行成功	作业在计算集群中被成功的执行并执行完毕，单击作业列表中相应的日志按钮可查看相关日志
运行失败	作业在计算集群中执行出现异常，单击作业列表中相应的日志按钮可查看相关日志

1.6 E-MapReduce常用文件路径

E-MapReduce 是阿里云上的开源大数据平台，用户可以登录集群主节点查看相关安装路径。登录后也可以使用 `env |grep xxx` 查看。

大数据组件目录

软件安装目录在 `/usr/lib/xxx` 下，例如：

- Hadoop: `/usr/lib/hadoop-current`
- Spark: `/usr/lib/spark-current`
- Hive: `/usr/lib/hive-current`
- Flink: `/usr/lib/flink-current`

日志目录

组件日志目录在 `/mnt/disk1/log/xxx` 下，例如：

- Yarn ResourceManager 日志: master 节点 `/mnt/disk1/log/hadoop-yarn`
- Yarn NodeNanager 日志: slave 节点 `/mnt/disk1/log/hadoop-yarn`
- HDFS NameNode 日志: master 节点 `/mnt/disk1/log/hadoop-hdfs`
- HDFS DataNode 日志: slave 节点 `/mnt/disk1/log/hadoop-yarn`
- Hive 日志: Master 节点 `/mnt/disk1/log/hive`

配置文件

配置文件目录在 `/etc/ecm/xxx` 下，如果仅登录节点修改集群中的配置文件，配置文件不会生效。



注意:

此处的配置文件只能浏览配置文件的参数，如需修改参数，请到 EMR 控制台操作。

- **Hadoop:** `/etc/ecm/hadoop-conf/`
- **Spark:** `/etc/ecm/spark-conf/`
- **Hive:** `/etc/ecm/hive-conf/`